Outline

L Lathrop

Sentiment analysis of presidential campaign speeches using R

Data sources

This data project makes use of three speeches each for the remaining three candidates, as follows:

Democratic candidates

Hillary Clinton
- Speech transcript at AIPAC Policy Conference (03/21/16)
- Remarks at Suffolk County Democratic Committee's Annual Spring Dinner (04/12/16)
- Remarks at Primary Rally in Philadelphia (04/26/16)

Bernie Sanders
- Remarks on Wall Street and the Economy in New York City (01/05/16)
- Remarks at the New Hampshire Democratic Party Jefferson-Jackson Dinner in Manchester (11/29/15)
- Remarks on Democratic Socialism in the United States (11/19/15)

Republican candidate

Donald Trump
- Speech at AIPAC Policy Conference (03/21/16)
- Foreign policy speech (04/27/16)
- Remarks Announcing Candidacy for President in New York City (06/16/15)

Introduction

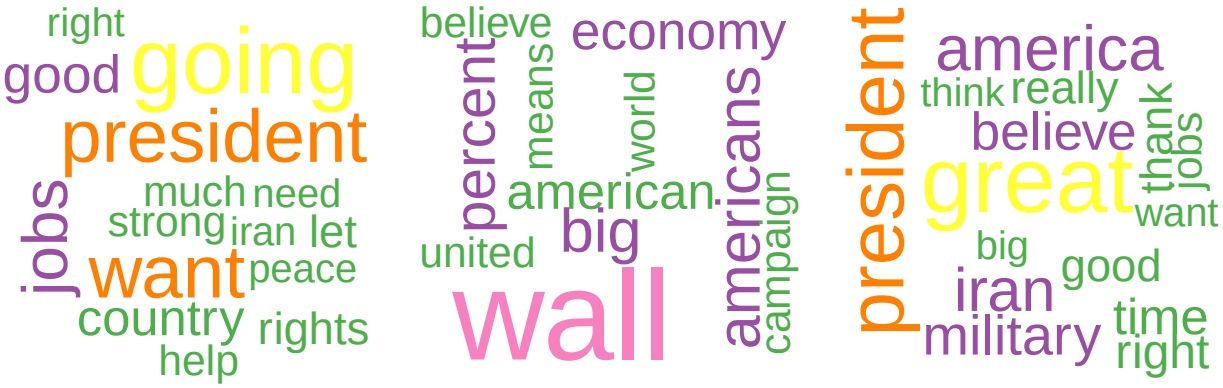
1. Brief introduction to text mining, sentiment analysis, and topic modeling. *(Sources: Chen, et al, Exploiting domain knowledge in aspect extraction; Liu, Sentiment analysis and opinion mining; Liu, Sentiment analysis and subjectivity.)*

2. Introduction to this analysis of political speech and why it is both relevant and important.

Section 1: Basic text analysis

1. Description of speeches being used.
2. Basic description of text processing.
3. Most common words used by each candidate as shown in wordclouds (N.B. These wordclouds are VERY rudimentary and generic, both in terms of colors and the style of the clouds. In other words, the final graphics will look very different. I'm planning to use visualization of the kind seen here.)

4. Brief analysis of results.



Section 2: Sentiment analysis

1. Description of processing.

2. Sentence-based sentiment analysis by comparing individual statements against the `pros and cons dataset` used in (Ganapathibhotla and Liu, Coling-2008 & Liu, Hu and Cheng, WWW-2005) for determining context (aspect) dependent sentiment words, which are then applied to sentiment analysis of comparative sentiences. Below are preliminary visualizations for this analysis.

3. Analysis of results including percentages of highly positive and highly negative scores for each candidate. For example, Hillary Clinton's sentiments scores reveal that 40.16% of her statements are highly positive and 14.17% are highly negative.

Section 3: Topic models using Latent Dirichlet Allocation (LDA)

1. Brief description of topic modeling, how it works, and why it is used
2. Brief description of processing used, e.g., log-liklihood

3. I'm not certain about the visualization I will use here, but it will likely be some sort of network or cluster visualization showing the links between topics, such as this one

Concluding summary